Biophysical Journal Volume 104 January 2013 453–462

Absolute Hydration Free Energies of Blocked Amino Acids: Implications for Protein Solvation and Stability

Gerhard König,[†*] Stefan Bruckner,[‡] and Stefan Boresch[‡]

[†]Laboratory of Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland; and
[‡]Department of Computational Biological Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria

ABSTRACT Most proteins perform their function in aqueous solution. The interactions with water determine the stability of proteins and the desolvation costs of ligand binding or membrane insertion. However, because of experimental restrictions, absolute solvation free energies of proteins or amino acids are not available. Instead, solvation free energies are estimated based on side chain analog data. This approach implies that the contributions to free energy differences are additive, and it has often been employed for estimating folding or binding free energies. However, it is not clear how much the additivity assumption affects the reliability of the resulting data. Here, we use molecular dynamics–based free energy simulations to calculate absolute hydration free energies for 15 N-acetyl-methylamide amino acids with neutral side chains. By comparing our results with solvation free energies for side chain analogs, we demonstrate that estimates of solvation free energies of full amino acids based on group-additive methods are systematically too negative and completely overestimate the hydrophobicity of glycine. The largest deviation of additive protocols using side chain analog data was 6.7 kcal/mol; on average, the deviation was 4 kcal/mol. We briefly discuss a simple way to alleviate the errors incurred by using side chain analog data and point out the implications of our findings for the field of biophysics and implicit solvent models. To support our results and conclusions, we calculate relative protein stabilities for selected point mutations, yielding a root-mean-square deviation from experimental results of 0.8 kcal/mol.

INTRODUCTION

When facing complex phenomena, scientists are often forced to rely on approximations. In the context of biophysics, one of the most widely employed assumptions is that of additivity, i.e., that the free energy of a molecule can be estimated from the free energies of its fragments. Thermodynamic additivity has been referred to as the "fourth law of thermodynamics" by Benson (1). However, it is difficult to assess the correctness of this assumption—especially for macromolecules. On the one hand, exact thermodynamic data on large compounds such as proteins is sparse and very difficult to obtain. On the other hand, seemingly additive behavior can be the result of fortuitous error-compensation, which arises from the interplay of a multitude of degrees of freedom present in biomolecular systems. Examples for error compensation include the enthalpy-entropy compensation in aqueous solution (2) or the error compensation due to the use of thermodynamic cycles (3). Because those terms depend on the environment of the system, the resulting errors are system-dependent. For example, the data on the inaccuracy of binding free energy predictions will not be directly transferable to studies of the free energy of folding.

One common feature of all biological processes is the presence of aqueous solvent, which causes the hydrophobic effect (4). In proteins, water influences a wide spectrum of processes, including folding (5,6), stability (7), and dynamics (8). Furthermore, water is one of the main actors in ligand binding (9) and for the selectivity of biochemical interactions (10). From a biophysical point of view, it is, therefore, essential to take into account the functional role of the solvent for complex environments, such as biomolecules. This means that no prediction of a biological process can be accurate if the prediction of the contribution of water is wrong—making the prediction of solute-solvent interactions (e.g., in form of solvation free energies) one of the most important and general benchmark tests in biomolecular simulation (11).

Unfortunately, solvation free energies suffer from the same problem as most other energetic studies in biomolecules: Until now it has not been possible to measure the solvation free energy of proteins or even amino acids experimentally (12). Therefore, estimates of the solvation free energy have relied on the additivity assumption by first measuring solvation free energies of small model compounds considered representative of parts of the system of interest, followed by adding the contributions of the model compounds. In particular, full amino acids were conceptually split into N-methylacetamide, representing the backbone (13), and so-called side chain analogs, e.g., methanol for Ser, etc. (14). These data or rather estimates of the solvent affinity of individual amino acids are in turn used to characterize peptides or proteins, again assuming an essentially additive relationship between individual contributions.

Side chain analog data is one component in many hydrophobicity scales (15). E.g., the still widely used hydropathy

Submitted June 18, 2012, and accepted for publication December 10, 2012.

*Correspondence: gerhard.koenig@nih.gov

Editor: Bert de Groot.

0006-3495/13/01/0453/10 $2.00

http://dx.doi.org/10.1016/j.bpj.2012.12.008

index of Kyte and Doolittle (16) was derived from a combination of water-vapor transfer free energies of side chain analogs, i.e., essentially the data reported in Wolfenden et al. (14), and the interior-exterior distribution of amino-acid side chains determined by Chothia, refined by modifications based on chemical intuition (16). Small model compound data such as side chain analog transfer free energies also form the foundation of fragment-based methods to estimate partition coefficients log P (17–20) or solvation free energies (21) of nonionic organic solutes. To a lesser degree, it also applies to certain solvent-accessible-surface-based implicit solvent models.

However, for free energy differences and entropies, there is no physical basis for the additivity assumption, and this has led to some (relatively unheeded) criticism. Several studies pointed out inconsistencies between various hydrophobicity scales (22). Lazaridis et al. (23) demonstrated in an elegant thought-experiment that the group additivity approximation breaks down for polar and charged groups. Avbelj and Baldwin (24) showed that group additivity does not hold for the peptide group. Della Gatta et al. (25) reported solvation enthalpies for peptides and dipeptides; their measurements also provide clear experimental evidence that solvation enthalpies of the peptide groups in the backbone are not additive. Various aspects of the energetics of protein folding, including an analysis of the failure of the principle of group additivity for the polar NH and CO groups, were discussed by Baldwin (26). In 1997, Robertson and Murphy (27) considered the nonadditivity of energetic contributions from the various groups that make up polar and nonpolar surfaces to be the number-one culprit for the deviation of 57–182% between calculated and experimental results for the ΔC_p of unfolding. However, concerning the supposed nonadditivity, they lamented that no straightforward approach is available yet for evaluating its role.

Group additivity can be viewed as a special case of context dependence; in this case, the underlying question becomes how introducing a functional group X (e.g., a methyl or hydroxyl moiety) changes the free energy of interest (e.g., the solvation free energy of the full molecule) as a function of the scaffold to which it is attached. One theoretical framework to describe such effects would be the conditional free energies of Ben-Naim and co-workers (28–30); by a careful analysis of experimental data, they have been able to tabulate conditional free energy contributions for several functional groups relevant to protein stability.

In a review on the subject of additivity, Dill (31) provides a balanced discussion of the pros and cons of the additivity principle. He argues pragmatically that while there may be no theoretical justification, the assumption of additivity is useful provided it is sufficiently accurate. Concerning applications to protein folding, he points out that free energy differences between the folded and unfolded states are typically <10 kcal/mol. Thus, he suggests one-tenth of this upper bound to be acceptable as error; i.e., the total error resulting from an approach based on the group additivity assumption should be <1 kcal/mol. For a protein consisting of 100 residues, this means that the error per residue must not exceed $1/\sqrt{100} = 0.1$ kcal/mol, assuming random errors only. If, however, the errors were systematic, the maximum allowable error per residue is ~0.01 kcal/mol.

We recently computed relative solvation free energy differences for several pairs of N-acetyl-methylamide amino acids, such as Ala → Ser, and compared them with the corresponding results for side chain analogs, such as methane → methanol (32). We found differences between side chain analog and amino-acid solvation free energy differences of up to 66% (or, in absolute numbers, 4.9 kcal/mol) for the pair Ala-Ser, and could discern two sources for this observed nonadditivity:

The first is a steric effect, to which we refer as "solvent exclusion." It accounts for the fact that atoms in the interior of a large molecule do not interact with water and, therefore, will not contribute to the solvent affinity of the molecule. Several techniques account for solvent exclusion by scaling the solvation free energy contribution of an atom or fragment by its solvent accessible surface area (33–39), and also some implicit solvent models rely on this approximation (37,38,40–42), often in conjunction with the side chain analog data by Wolfenden et al. (14).

The second, even larger contribution, referred to as "self-solvation" (43,44), was identified after 'solvent exclusion' was found insufficient for explaining the observed differences. We showed that, especially in the gas phase, polar side chains stabilize themselves by interacting with the backbone, thus considerably reducing the desolvation penalty compared to what one would expect from side chain analog data. In related work, Chang et al. (45) published a comparison of hydration free energies of zwitterionic and nonzwitterionic amino acids, as well as their corresponding side chain analogs. Because the self-solvation effect was not very pronounced for the two selected systems, their results correlated reasonably well with the side chain analog data.

In this study, we assessed the errors that arise due to the additivity hypothesis for absolute solvation free energies of N-acetyl-methylamide amino acids. The data provide a missing link between earlier computational work (32,45) and experiment (14). By having available absolute solvation free energy differences, a much more direct comparison to the experimental data is possible. In contrast to the pure amino acids used by Chang et al. (45), the blocking groups add peptide bonds to the two ends of the amino acid, thus resolving two problems: While the zwitterionic amino acids of Chang et al. (45) are representative for the situation found in solution, one is rather unlikely to encounter the zwitterionic form in the gas phase; the opposite is true for neutral amino acids, which reflect the most likely state in the gas phase, but not in solution.

By focusing on amino acids, our simulations did not include effects of tertiary structure. Because nonbonded interactions in tertiary structure contacts are likely to increase the errors of the additivity hypothesis, our results thus represent the best-case scenario for an additivity-based approach that is applied to proteins. On the other hand, the presence of the blocking groups makes the solutes resemble peptides, with the core of a peptide backbone present. Thus, the simulation results can account for possible interactions between the side chain and its backbone, i.e., self-solvation. Such interactions represent the major contribution to the error on the level of the primary structure. In addition, the quasi-peptide bond makes it possible to study the effect of the secondary structure on peptide solvation. Because group-based additive approaches do not account for changes of the structure, any dependencies of the solvation free energy on the secondary structure will directly add to the error due to the additivity hypothesis. Comparing the computational results of this work to the side chain analog data by Wolfenden et al. (14), we focus on two questions:

1. What is the magnitude of potential errors when applying additivity principles to the computation of solvation free energies of amino acids by relying on side chain analog data?
2. Are the associated errors systematic or random?

The remainder of this article is organized as follows: Details of the model systems and simulations are summarized. We then present the results for the absolute solvation free energies of blocked amino acids and compare them with the results of the corresponding side chain analogs. To lend support to our data, we calculate the effect of solvation on unfolding free energies of selected point mutations. We conclude with a short discussion of our findings and their biological relevance. A comparison of our explicit solvent results with several implicit solvent models is shown in the Supporting Material.

METHODS

We calculated the solvation free energies of all canonical neutral N-acetyl-X-methylamide amino acids (X = Ala, Val, Leu, Ile, Ser, Thr, Cys, Met, Asn, Gln, His, Phe, Tyr, Trp). We did not attempt analogous simulations for the charged amino acids (Arg, Asp, Lys, Glu) because they require complex corrections for the finite-range treatment of electrostatic interactions (46,47). We also omitted the amino-acid Proline because there is no corresponding side chain analog. However, the two tautomeric states of neutral histidine were considered (referred to as Hid and Hie, with the proton attached to the δ- and ϵ-nitrogens, respectively). To save computational costs, we began by calculating relative solvation free energy differences of all amino acids with respect to an intermediate state, which we denote as pseudo-glycine (PG). It resembles Gly, except for the atom type of the C_α carbon. (Note that atom type CT1 instead of CT2 was used for the C_α carbon. This has small effects on Lennard-Jones interactions, the cross-term map potential, and dihedral terms involving the C_α carbon. The partial atomic charges were those of glycine.) We then calculated the absolute solvation free energy of PG. Thus, the absolute solvation free energy of each amino acid is the sum of one absolute and one relative solvation free energy difference, i.e.,

$$\Delta A^{\mathrm{solv}}_{\mathrm{aa}} = \Delta A^{\mathrm{solv}}_{\mathrm{PG}} + \Delta\Delta A^{\mathrm{solv}}_{\mathrm{PG}\rightarrow\mathrm{aa}}. \quad (1)$$

Each relative solvation free energy difference was calculated with the familiar thermodynamic cycle (48) involving the alchemical mutation of the amino acid under consideration to PG in the gas phase and in aqueous solution. Because electrostatic and Lennard-Jones interactions were modified separately, the calculation of $\Delta\Delta A^{\mathrm{solv}}_{\mathrm{PG}\rightarrow\mathrm{aa}}$ consisted of four steps:

Step 1. The charges of the side chain were turned off in explicit solvent ($\Delta A^{\mathrm{H_2O}}_{\mathrm{aa}\rightarrow\mathrm{unch\cdot aa}}$).
Step 2. The uncharged side chain was mutated to PG ($\Delta A^{\mathrm{H_2O}}_{\mathrm{unch\cdot aa}\rightarrow\mathrm{PG}}$).
Steps 3. and 4. The analogous steps were carried out in the gas phase ($\Delta A^{\mathrm{gas}}_{\mathrm{aa}\rightarrow\mathrm{unch\cdot aa}}$ and $\Delta A^{\mathrm{gas}}_{\mathrm{unch\cdot aa}\rightarrow\mathrm{PG}}$). Thus, each relative solvation free energy was calculated according to

$$\begin{aligned}\Delta\Delta A^{\mathrm{solv}}_{\mathrm{PG}\rightarrow\mathrm{aa}} = &-\Delta A^{\mathrm{H_2O}}_{\mathrm{aa}\rightarrow\mathrm{unch\cdot aa}} - \Delta A^{\mathrm{H_2O}}_{\mathrm{unch\cdot aa}\rightarrow\mathrm{PG}} \\ &+ \Delta A^{\mathrm{gas}}_{\mathrm{aa}\rightarrow\mathrm{unch\cdot aa}} + \Delta A^{\mathrm{gas}}_{\mathrm{unch\cdot aa}\rightarrow\mathrm{PG}}.\end{aligned} \quad (2)$$

Then the absolute solvation free energy of PG was computed in three additional steps:

Step 5. The partial charges of PG in solution were turned off ($\Delta A^{\mathrm{H_2O}}_{\mathrm{PG,unch}}$).
Step 6. The Lennard-Jones interactions between PG and water ($\Delta A^{\mathrm{H_2O}}_{\mathrm{PG,vdw}}$) were switched off.
Step 7. Because of the technical realization of alchemical mutations in CHARMM (49,50), Steps 5 and 6 also remove all intrasolute nonbonded interactions; this loss of interactions was accounted for by a gas-phase correction ($\Delta A^{\mathrm{gas}}_{\mathrm{PG}}$), restoring all intrasolute nonbonded interactions. Thus, the absolute solvation free energy of PG is given by

$$\Delta A^{\mathrm{solv}}_{\mathrm{PG}} = -\Delta A^{\mathrm{H_2O}}_{\mathrm{PG,unch}} - \Delta A^{\mathrm{H_2O}}_{\mathrm{PG,vdw}} + \Delta A^{\mathrm{gas}}_{\mathrm{PG}}. \quad (3)$$

All free energy calculations were conducted with CHARMM (49,50), using the CHARMM27 force field that includes the backbone cross-term map correction (51,52). Most free energy differences were computed with Bennett's acceptance ratio method (53). Some of the gas phase corrections, $\Delta A^{\mathrm{gas}}_{\mathrm{unch\cdot aa}\rightarrow\mathrm{PG}}$ and $\Delta A^{\mathrm{gas}}_{\mathrm{PG}}$, were computed by thermodynamic integration (54). Table 1 provides an overview of all simulations carried out. In particular, we list the respective number of λ-states (*second column*), simulation times per λ-state (*third column*), the total simulation length for

TABLE 1 Overview of the simulation protocols

Type of mutation	No. of λ[a]	ns/λ[b]	Total time (ns)	Method[c]
$\Delta A^{\mathrm{H_2O}}_{\mathrm{aa}\rightarrow\mathrm{unch\cdot aa}}$	3	10	30	NBB
$\Delta A^{\mathrm{H_2O}}_{\mathrm{unch\cdot aa}\rightarrow\mathrm{PG}}$	5–7	10	50–70	BAR
$\Delta A^{\mathrm{gas}}_{\mathrm{aa}\rightarrow\mathrm{unch\cdot aa}}$	3	84	252	TI
$\Delta A^{\mathrm{gas}}_{\mathrm{unch\cdot aa}\rightarrow\mathrm{PG}}$	21	4	84	TI
$\Delta A^{\mathrm{protein}}_{\mathrm{aa}\rightarrow\mathrm{PG}}$	11	5	55	TI
$\Delta A^{\mathrm{H_2O}}_{\mathrm{PG,unch}}$	3	10	30	BAR
$\Delta A^{\mathrm{H_2O}}_{\mathrm{PG,vdw}}$	9	10	90	BAR
$\Delta A^{\mathrm{gas}}_{\mathrm{PG}}$	21	4	84	BAR

[a]Number of λ-states, including the end points $\lambda = 0$ and $\lambda = 1$.
[b]Simulation time per λ-state.
[c]Free energy method employed for the mutation: NBB, non-Boltzmann Bennett (61); BAR, Bennett's acceptance ratio method (53); TI, thermodynamic integration (54).

each type of calculation (*fourth column*), and the free energy method (*fifth column*). Each free energy difference was computed independently four times by starting the simulations with different initial random velocities. The standard deviations of the four values for each free energy difference of interest were combined by Gaussian error propagation; the statistical error estimate obtained in this manner is reported in the rightmost column of Table 2.

Gas-phase free energy differences were calculated using Langevin dynamics simulations with a friction coefficient of 5 ps^{-1} on all atoms. Random forces were applied according to the target temperature of 300 K, and hydrogen masses were set to 10 atomic mass units (amu) to justify a time step of 2 fs. For analysis, trajectories were written every 100 steps.

In all solvent simulations, 862 TIP3P water molecules (55,56) were present. The simulation box was a truncated octahedron. The side length L of the cube from which the octahedron was generated was L = 37.25 Å, which was the average box size over all selected amino acids (the optimal box size of each amino acid was determined from 1-ns constant pressure simulations). For the determination of $\Delta A^{H_2O}_{PG,vdw}$, we used constant pressure simulations. Integration of the equations of motion was carried out with the velocity-Verlet algorithm as implemented in the TPCNTRL module of CHARMM (57); the time step was 2 fs. The temperature was maintained at ~300 K using separate Nosé-Hoover thermostats (58) for solute and solvent. SHAKE (59) was used to keep the water geometry rigid. Lennard-Jones interactions were switched off between 10 and 12 Å, while electrostatic interactions were computed with the particle-mesh Ewald method (60). Coordinates obtained after 1 ns of equilibration served as the starting configuration for the free energy simulation. In addition, each system was equilibrated for 100 ps at every λ-value.

To overcome slow sampling of side chain rotamers when computing $\Delta A^{H_2O}_{aa\rightarrow unch\cdot aa}$ and $\Delta A^{gas}_{aa\rightarrow unch\ aa}$, we lowered the energy barriers of the side chain torsional angles χ_1 and χ_2 by deleting the corresponding dihedral angle terms. To obtain correct free energies the data were reweighted with the non-Boltzmann Bennett method (61) according to the value of the dihedral potential.

To test the accuracy of the solvation free energies, we conducted free energy simulations of point mutations in three different proteins: A130S of T4 lysozyme (PDB:118L), A104C of rat intestinal fatty acid-binding protein (PDB:1IFC), and F22Y of bovine pancreatic phospholipase A2 (PDB:1UNE). The initial structures were prepared with CHARMM-GUI (62), adding a solvation shell of 13 Å and KCl to achieve a 150 mM solution. The simulation boxes were truncated octahedra, with L = 81.725, 71.8535, and 74.906 Å, respectively. Coordinates obtained after 1 ns of equilibration at constant pressure served as the starting configuration for the free energy simulations. The proteins were simulated with the same energy settings as the amino acids. For each mutation, the free energy difference to PG was computed using 11 λ-steps (changing electrostatic and Lennard-Jones interactions simultaneously), leading to $\Delta A^{protein}_{aa\rightarrow PG}$. To avoid double-counting the changes of interactions within the amino acid, the gas-phase correction terms have to be incorporated to form

$$\Delta\Delta A^{protein}_{PG\rightarrow aa} = \Delta A^{gas}_{aa\rightarrow unch.aa} + \Delta A^{gas}_{unch.aa\rightarrow PG} - \Delta A^{protein}_{aa\rightarrow PG}, \quad (4)$$

which can be regarded as a protein solvation free energy difference (i.e., treating the protein as a solvent). This step allows the direct comparison with solvation free energies and, ultimately, the calculation of protein stabilities in combination with the solvation free energies of amino acids. The free energy difference associated with a point mutation that turns amino acid X into Y is given by

$$\Delta\Delta A^{mut}_{protein} = \Delta\Delta A^{protein}_{PG\rightarrow Y} - \Delta\Delta A^{protein}_{PG\rightarrow X}. \quad (5)$$

TABLE 2 Solvation free energies of blocked amino acids (in kcal/mol) and the contributions from gas phase and solution to the mutation to pseudo-glycine

Mutation	$\Delta A^{H_2O}_{PG\rightarrow aa}$ [a]	$\Delta A^{gas}_{PG\rightarrow aa}$ [b]	$\Delta\Delta A^{solv}_{PG\rightarrow aa}$ [c]	ΔA^{solv}_{aa} [d]	σ [e]
Ala	14.0	11.1	2.9	**−12.0**	0.1
Asn	−65.7	−63.1	−2.6	**−17.5**	0.3
Cys	14.9	13.1	1.8	**−13.1**	0.2
Gln	−43.5	−40.5	−3.0	**−17.9**	0.1
Gly	0.8	0.4	0.4	**−14.5**	0.1
Hid	−8.2	−2.0	−6.2	**−21.1**	0.2
Hie	−23.0	−19.6	−3.3	**−18.2**	0.2
Ile	21.1	17.1	4.0	**−10.9**	0.5
Leu	2.4	−1.2	3.6	**−11.3**	0.6
Met	12.7	10.3	2.4	**−12.5**	0.2
Phe	21.4	18.8	2.6	**−12.3**	0.4
Ser	18.5	18.4	0.1	**−14.8**	0.3
Thr	−2.7	−4.0	1.3	**−13.6**	0.5
Trp	21.0	21.3	−0.3	**−15.2**	0.2
Tyr	4.2	6.2	−2.0	**−16.9**	0.1
Val	15.2	11.8	3.3	**−11.6**	0.2

[a]Relative free energies to pseudo-glycine (PG) in aqueous solution $\left(\Delta A^{H_2O}_{PG\rightarrow aa} = -\Delta A^{H_2O}_{aa\rightarrow unch\cdot aa} - \Delta A^{H_2O}_{unch\cdot aa\rightarrow PG}\right)$.
[b]Relative free energies to PG in gas phase $\left(\Delta A^{gas}_{PG\rightarrow aa} = -\Delta A^{gas}_{aa\rightarrow unch\cdot aa} - \Delta A^{gas}_{unch\cdot aa\rightarrow PG}\right)$.
[c]Relative solvation free energies to PG (ΔA^{solv}_{PG} = −14.9 kcal/mol).
[d]Absolute solvation free energies.
[e]Standard deviations of ΔA^{solv}_{aa}.

RESULTS AND DISCUSSION

Accuracy of calculated free energy differences

The absolute solvation free energy differences for 15 blocked amino acids (Gly, Ala, Val, Leu, Ile, Ser, Thr, Cys, Met, Asn, Gln, Phe, Tyr, Trp, and the two uncharged His tautomers Hid and Hie; see Methods), using the CHARMM27 (51,52) force field, are reported in Table 2. Because no experimental data are available, it is important to obtain a good understanding of the size of the error that may be present in our results. Calculations of the solvation free energies of side chain analogs and other small compounds often serve as a gauge for the accuracy of force fields (46,63–67); hence, extensive data is available on the error margins of such free energy simulations. Shirts et al. (46) obtained a root-mean-square deviation (RMSD) of 1.3 kcal/mol for all uncharged amino-acid side chain analogs using the CHARMM force field. They also employed the AMBER and OPLS-AA force fields, with very similar agreement to experiment. In a blind test for 17 small molecules, various methods to compute free energy differences yielded RMS errors between 1.3 and 1.7 kcal/mol (12). In the largest such study reported to date (hydration free energies of 504 neutral small organic molecules), the RMS error over the whole set was 1.3 kcal/mol (67). It seems justified to assume that our error margins will be comparable to the deviations found in these studies.

We did not recalculate the solvation free energy differences for the side chain analogs because extremely precise data are available in the work by Shirts et al. (46). Relative free energy differences between pairs of side chain analogs were reported in our earlier study (32) and agreed extremely

well with values calculated from Shirts et al. (46). In addition, absolute solvation free energy differences for ethane, methanol, and 3-methyl-indole were computed and found to agree within ± 0.3 kcal/mole with the data reported in Shirts et al. (46) (and C. Rauer and S. Boresch, unpublished). Similarly, relative solvation free energy differences between pairs of blocked amino acids reported earlier (32) agree quite well with the data we present here.

Comparison to data for side chain analogs

In Fig. 1, we compare the solvation free energy results for N-acetyl-methylamide amino acids (+) from our study with the results for the side chain analog results (×) by Shirts et al. (46) obtained with the CHARMM force field. The computational results (ordinate) are plotted against the experimental data by Wolfenden et al. (14) (abscissa). Because the aim of this study is the assessment of the additivity hypothesis for side chain analog solvation free energies, we show all results relative to the respective Gly reference state, i.e., a blocked Gly for the blocked amino acids and H_2 in the case of side chain analogs. The one-letter codes of the amino acids locate their horizontal position in the plot.

We also include regression lines for the two data sets. If solvation free energies were truly additive, the two lines should be identical and, in an ideal setting, form a perfect diagonal; i.e., for a regression line $y = kx + d$, we should find a slope $k = 1.0$ and an axis intercept $d = 0$. However,

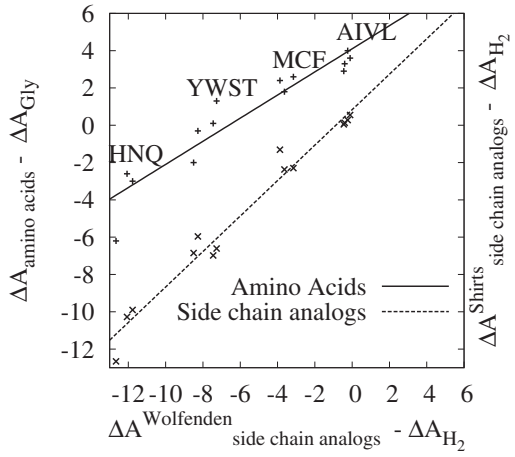

FIGURE 1 Comparison of blocked amino-acid solvation free energies with their corresponding side chain analog data (kcal/mol). (*Amino acids*) Results for the blocked amino acids reported in Table 2 relative to Glycine. (*Side chain analogs*) Side chain analog results of Shirts et al. (46) relative to the side chain analog of Glycine (H_2). The approximate horizontal position of each amino acid is indicated by its one-letter code. The computational results relative to Gly ($\Delta A_{\text{amino acids}} - \Delta A_{\text{Gly}}$) or its side chain analog H_2 ($\Delta A^{\text{Shirts}}_{\text{side chain analogs}} - \Delta A_{H_2}$; y axis) are plotted against the experimental solvation free energy differences of the side chain analogs reported by Wolfenden et al. (14), also relative to H_2 ($\Delta A^{\text{Wolfenden}}_{\text{side chain analogs}} - \Delta A_{H_2}$; x axis). If the contributions of side chain and backbone to the solvation free energy were purely additive, both lines should be identical.

as can be seen in Fig. 1, both the slopes and the axis intercepts of the two regression lines deviate from these ideal values. The slope closest to 1.0 was found for the side chain analogs ($k = 0.95$), whereas the slope of the blocked amino acids is only $k = 0.62$. Because the abscissa denotes experimental solvation free energies for the side chain analogs while the ordinate shows computational results, the small deviation of 0.05 of the slope of the side chain analog data from unity can be attributed to imperfections of the force field. Such deviations can be expected for all regression lines, so this value gives us an idea of the acceptable incongruities between the slopes; i.e., if the slopes of the two regression lines differ by $\gg 0.05$, then they are unlikely to be caused by errors of the force field.

The slopes in Fig. 1 are a measure of how adding a functional group (e.g., a hydroxyl group) to a molecule changes the solvation free energy of the resulting compound. If the slope is steep, adding the functional group to the molecule will change the solvation free energy drastically. On the other hand, if the slope were completely flat, the solvent affinity of the molecule would not be affected by the addition of the functional group. The different slopes in Fig. 1 clearly demonstrate that the solvation free energy difference associated with the addition of a functional group depends on the scaffold to which it is attached. For example, adding a hydroxyl group to methane (the side-chain analog of alanine) changes the solvation free energy by 7.0 kcal/mol. However, for a blocked alanine, the change in solvation free energy from adding a hydroxyl group is only 2.8 kcal/mol. This reduction of the relative solvation free energy differences corresponds exactly to what one would expect if interactions between side chain and backbone weaken the solvent affinity of the hydroxyl group. The solvation free energy results presented here are in perfect agreement with the findings and analysis of relative mutations between amino acids in a previous publication (32).

Another interesting aspect in Fig. 1 are the different axis intercepts of the regression lines. These can be explained by the relative ranking of Gly in the respective list of solvation free energies. Because the solvation free energy difference of Gly, or its side chain analog H_2, was subtracted from all data points in the plot, this has a tremendous effect on the origin of the regression line. Though the relative ordering of most amino acids is changed only by one rank when comparing side chain analogs and blocked amino acids, the position of Gly differs dramatically: In our results for the blocked amino acid, Gly is more hydrophilic than 8 out of 15 amino acids. However, according to the side chain analog data, the analog of Gly (H_2) is the most hydrophobic compound. Although the relative insolubility of such a volatile gas in water stands to reason, inferring the same for Gly is rather counterintuitive. The relative ranking of Gly already highlights how treacherous the employment of additivity principles may be.

When comparing the absolute solvation free energies of side chain analogs by Wolfenden with the blocked amino-acid data (including glycine), a correlation coefficient R^2 of 0.82 is found. This indicates that, to a certain extent, amino-acid solvation free energies can be estimated from side chain analog data. A linear least-square fit of the two data sets leads to

$$\Delta A_{aa}^{solv} \approx 0.49\ \Delta A_{sc}^{solv} - 12.73, \quad (6)$$

where ΔA_{sc}^{solv} is the side chain analog solvation free energy in kcal/mol. Compared to the calculated ΔA_{aa}^{solv} in Table 2, the above equation yields an RMSD of 1.0 kcal/mol, which is a significant improvement over the RMSD of 6.4 kcal/mol seen in Fig. 1. For many applications of side chain analog data, the scaling factor of 0.49 might be used to crudely account for the interactions between side chain and backbone. However, this correction is only admissible for relative comparisons of amino acids (i.e., $\Delta\Delta A^{solv}$) and if the effects from secondary and tertiary structure are assumed to be small (or to cancel out). To a certain extent, the correction might also be used to scale the parameters of implicit solvent models that were derived from side chain analog data. However, the resulting parameters would have to be tested very carefully, because using this correction implies that all atom types contribute equally to nonadditivity.

The different levels of context dependence in proteins

At this point it is instructive to look at some numerical examples of the nonadditivity found in our calculations. In amino acids, the context-dependence can be considered at three different levels.

For the first level, we compare the effect of adding a chemical group (e.g., CH_3) to the side chain. Table 3 lists the change of solvation free energy associated with adding methyl, hydroxyl and thiol groups to amino acids (*third column*) and their corresponding side chain analogs (last column). When adding an apolar methyl group, the $\Delta\Delta A_{aa}^{solv}$ ranges from 0.7 to 2.5 kcal/mol for the amino acids, while the side chain analog results range between −0.5 and 0.2 kcal/mol. This highlights that adding methyl groups to amino acids has a considerably more hydrophobic effect than expected from side chain analog data. The increased hydrophobicity can be explained in terms of solvent exclusion, i.e., the additional methyl groups prevent water from interacting with the hydrophilic backbone atoms. Thus the favorable contribution of the backbone to the solvation free energy is partially lost. Our data shows that the naive use of unscaled side chain analog data can lead to RMSDs of 1.2 kcal/mol for the treatment of methyl groups. When adding the polar hydroxyl and thiol groups to the side chain, the RMSD becomes higher, reaching 1.7 kcal/mol. As shown in our previous study (32), solvent exclusion also plays a role when adding polar groups to the side chain, but the main contribution to nonadditivity originates from the self-solvation effect. Both hydroxyl and thiol groups are able to form hydrogen bonds with their own backbone. If such an intraresidue hydrogen bond is established, the favorable interaction energy with the solvent is lost, leading to increased hydrophobicity of the residue.

TABLE 3 Solvation free energy change associated with adding chemical groups (in kcal/mol)

Group	Mutation	ΔA_{aa}^{solv} [a]	ΔA_{sc}^{solv} [b]
$+CH_3$	Gly → Ala	+2.5	−0.5
	Ser → Thr	+1.2	+0.2
	Val → Ile	+0.7	+0.2
+OH	Ala → Ser	−2.8	−7.0
	Val → Thr	−2.0	−6.9
	Phe → Tyr	−4.6	−5.4
+SH	Ala → Cys	−1.1	−3.2

[a]Solvation free energy change based on blocked amino-acid data.
[b]Solvation free energy change based on side chain analog data.

The second level of context dependence in amino acids happens when attaching the side chain to the backbone. We use a naive fragment-based approach and estimate $\Delta\Delta A_{aa}^{solv}$ from the computed solvent affinity of Gly plus the experimental solvation free energy of the respective side chain analog. As already found in previous studies (32,45), the nonadditivity is strongest for polar amino acids. We will exemplify this with the most extreme case encountered in our work, Asn. The side chain analog of Asn is acetamide. Thus, by adding the solvation free energy of acetamide (9.7 kcal/mol (14)) to the solvation free energy of Gly (14.5 kcal/mol, this work), we obtain an estimated solvation free energy for Asn of 24.2 kcal/mol. This value overestimates the Asn result obtained in our study (17.5 kcal/mol) by almost 6.7 kcal/mol (or ~38%). For example, errors for Ser or Thr are not much smaller; they are 4.7 kcal/mol and 5.8 kcal/mol, respectively. The RMS error of such a naive fragment-based method over all amino acids, excluding Gly and using our Hid result for His, is ~4 kcal/mol. In addition, solvation free energies are always overestimated by the fragment-based approach, suggesting that the error is systematic.

The third level of context dependence involves the primary and secondary structure of a protein, i.e., the influence of the neighboring amino acids in the polypeptide chain. This level was studied by Staritzbichler et al. (68) by calculating solvation free energies of short peptides of varying length. Their results show that, up to a length of four residues, the contributions to the solvation free energy of each amino acid are almost independent, because there are no interactions between the backbone peptide bonds. As soon as the peptides are long enough to form secondary structures, the nonadditivity becomes more pronounced, leading to a significant reduction of the solvation free energy (~50% for non-alanine). This effect can be explained in terms of the self-solvation and solvent exclusion of the

backbone, and, theoretically, could be accounted for by tabulating solvation free energies for different combinations of secondary structures. On the level of tertiary structure, however, any hydrogen bond or salt bridge between side chains that are in close proximity to each other is going to decrease the solvation free energy of the protein. Such interactions cannot be accounted for a priori and, therefore, have to be calculated explicitly.

Comparison to protein denaturation free energies

Based on the discussion in the previous section, it might appear that additivity principles are of limited usefulness for applications in protein science. However, a notable exception is the unfolded state of proteins. By definition, the unfolded state does not contain stable secondary or tertiary structure components. Therefore, precomputed libraries of free energy differences of point mutations can be used to predict protein stability changes with relatively high accuracy (69). Blocked amino acids represent the simplest model for a particular amino acid in the unfolded state (70). In the unfolded state, the major interaction partner of each amino acid is water. The solvation free energy of the amino acid thus reflects its stability in the unfolded state. It is, therefore, possible to check relative solvation free energy results by linking them to the effect of point mutations on protein denaturation.

Starting from the assumption that the solvation free energy difference associated with a point mutation reflects the free energy difference in the unfolded state (i.e., $\Delta\Delta A_{AA}^{solv} \approx \Delta\Delta A_{AA}^{unfolded}$), the relative free energy of denaturation associated with the point mutation is given by

$$\Delta\Delta A_{protein}^{denat} = \Delta\Delta A^{solv} - \Delta\Delta A_{protein}^{mut}, \quad (7)$$

where $\Delta\Delta A_{protein}^{mut}$ is the free energy change associated with the mutation in the folded protein.

Under certain conditions, $\Delta\Delta A_{protein}^{mut}$ is zero and, therefore, the relative free energy of denaturation directly corresponds to the solvation free energy difference. This is the case if:

1. Both amino acids involved in the point mutation are solvent-exposed in the unfolded state and completely buried in a hydrophobic environment in the folded state. In an ideal setting, this would remove all solvent-solute interactions during the folding process. Thus, the denaturation free energy difference between the two endpoints would fully incorporate the solvation free energy difference of both amino acids.
2. Hydrogen bonds or salt bridges in the protein structure would introduce additional contributions to the protein stability. Therefore, the environment in the protein should be completely hydrophobic. This is often the case if, in the wild-type, the position in the protein was occupied by a buried hydrophobic amino acid. We will, therefore, only consider mutations that start from a buried hydrophobic amino acid.
3. The two amino acids involved in the mutation have to be isosteric, so that the apolar interactions with the environment are approximately equal and, therefore, their contribution will cancel out in the calculation of the relative denaturation free energy.

All of these conditions are met for the Val→Thr mutations in Fernández-Escamilla et al. (71) (the mutation sites are buried in a hydrophobic pocket and the RMSD between the mutated proteins is merely 0.28 Å). We will then assume that $\Delta\Delta A_{protein}^{mut}$ for this case is approximately zero. For cases where the above conditions are not met, $\Delta\Delta A_{protein}^{mut}$ has to be calculated explicitly. As a proof of concept, we have calculated $\Delta\Delta A_{protein}^{mut}$ for mutations of Ala→Ser, Ala→Cys, and Phe→Tyr in three different proteins. Table 4 lists the results of $\Delta\Delta A_{protein}^{mut}$ (*first column*) and the corresponding predictions of protein stability based on blocked amino-acid data ($\Delta\Delta A_{AA}^{denat}$, *second column*) and side chain analog data ($\Delta\Delta A_{SC}^{denat}$, *third column*). The experimental relative free energies of denaturation ($\Delta\Delta G_{protein}^{denat}$, *last column*) (71–74) were selected from the ProTherm database (75). A negative sign of $\Delta\Delta G_{protein}^{denat}$ indicates that the protein is destabilized by the mutation.

The comparison shows that side chain analog data is not able to explain the contribution of solvation to protein stability. In all cases, $\Delta\Delta G_{protein}^{denat}$ is significantly smaller than one would expect from side chain analog data, indicating that the expected deviations from additive behavior

TABLE 4 Comparison of relative denaturation free energy predictions based on relative solvation free energy differences of blocked amino acids and side chain analogs with experimental denaturation free energy results in kcal/mol

	$\Delta\Delta A_{protein}^{mut}$ [a]	$\Delta\Delta A_{AA}^{denat}$ [b]	$\Delta\Delta A_{SC}^{denat}$ [c]	$\Delta\Delta G_{protein}^{denat}$ [d]
Isosteric				
Val-Thr	—	−2.0	−6.9	−2.0 ± 1.2[e]
Quasi-isosteric				
Ala-Ser	−0.8 ± 0.6	−2.0	−6.2	−1.0 ± 0.9[f]
Ala-Cys	−3.5 ± 0.4	2.4	0.3	2.5 ± 0.5[g]
Phe-Tyr	−4.9 ± 1.1	0.3	−0.5	1.5 ± n/a[h]
RMSD[i]		**0.8**	**4.5**	

[a]Free energy difference of point mutation in protein (see Eq. 5).
[b]Predicted change of protein stability based on the difference of absolute solvation free energies ΔA_{aa}^{solv} of blocked amino acids from Table 2 (see Eq. 7).
[c]Predicted change of protein stability based on the difference of absolute solvation free energies of side chain analogs from Wolfenden et al. (14) (see Eq. 7).
[d]Experimental difference of free energies of denaturation for point mutations.
[e]Average of all six $\Delta\Delta G_{protein}^{denat}$ from Fernández-Escamilla et al. (71).
[f]Mutant A130S, from Blaber et al. (72).
[g]Mutant A104C, from Jiang and Frieden (73).
[h]Mutant F22Y, from Dupureur et al. (74).
[i]RMSD of the predicted $\Delta\Delta A_{X}^{denat}$ from experimental denaturation free energies $\Delta\Delta G_{protein}^{denat}$.

will probably be even higher in full proteins than for the solvation free energies of amino acids. This is reflected by the high RMSD of 4.5 kcal/mol from $\Delta\Delta G^{\text{denat}}_{\text{protein}}$ (presented in Table 4's *last row*).

The RMSD of 0.8 kcal/mol for the blocked amino-acid results, on the other hand, corresponds to the usual error range of free energy calculations for small molecules (46,65–67,76,77) and also agrees with the reported accuracy of computational predictions of protein stabilities conducted by Seeliger and de Groot (69). The denaturation free energy differences provide clear evidence that the reported nonadditivity does also occur in proteins.

FURTHER DISCUSSION

The data presented here are in accord with several studies that have criticized the usefulness of side chain analog data and related experimental hydrophobicity scales over the years. As early as 1981, Yunger and Cramer (43) suggested self-solvation as an important effect to consider when studying full amino acids instead of side chain analogs. The analysis presented in König and Boresch (32) clearly supports the interpretation of their results. The concept of self-solvation was pursued further by Roseman (44), who concluded in 1988 that absolute hydropathy scales must be determined experimentally from studies with peptides and polypeptides, rather than from studies with free amino acids and side chain analogs.

Our results show directly the magnitude of error incurred by the group additivity assumption. Given that the resulting error for blocked amino acids already ranges up to 6.7 kcal/mol, it is difficult to see how to meet Dill's criterion (31), i.e., an error of 1 kcal/mol or less for a full protein with additivity-based methods. In addition, because the hydrophilicity (especially of polar amino acids) is uniformly overestimated by the side chain analog data, the error is clearly biased (i.e., systematic). To return to our example of a hypothetical 100 amino-acid protein, this would mean an average error of 400 kcal/mol (not considering the errors of the secondary structure or tertiary structure contacts, which cannot be accounted for in a group-additive model). Such errors are clearly unacceptable. In all fairness it should also be noted that Dill's criterion indicates that the accuracy of force fields will have to improve as well, given RMSDs of ~1–2 kcal/mol between computed and measured solvation free energy differences of small molecules (46,65–67,76,77). However, modern force fields are well able to account for the effects of secondary and tertiary structure. In addition, their errors are, overall, mostly random, meaning that their application to our hypothetical protein would incur an error of ~10–20 kcal/mol only (which is ~20× better than with the group additivity-based approach). Thus free energy calculations based on molecular-dynamics simulations find themselves in a more auspicious initial position for future developments.

CONCLUSIONS

Based on the computed solvation free energy differences for 15 blocked amino acids, we find no justification for the assumption that the solvent affinity of amino acids can be described by adding individual contributions from backbone and side chain. Clearly, group additive models are not capable of capturing the complex interactions of amino acids in peptides and proteins with water on the one hand and neighboring residues on the other. Thus, they will also be inadequate for the description of more complex processes such as protein folding or ligand binding. This finding is further highlighted by the results of our protein stability calculations, which clearly demonstrate the superiority of the reported solvation free energies over side chain analog data.

To borrow an analogy from Wittgenstein (78), group-additivity-based methods should be regarded to be like a ladder that must be thrown away after one has climbed it. They have been useful in the past to understand the basic characteristics of biomolecules, but now have outlived their time. The correction factor for side chain analog data given in Eq. 6 can only be considered a very rough estimate for the backbone side chain interactions, and it does not account for effects of secondary or tertiary structure. Instead, given the availability of modern computer resources, more sophisticated methods should be applied. In the context of solvation free energies, the use of free energy simulations with explicit or generalized Born-based implicit solvent models (79,80) should be considered when dealing with structured biomolecules. To illustrate the quality of contemporary implicit solvent models, we include a comparison of our explicit solvent results with solvation free energies from three implicit solvent models in Table S2 of the Supporting Material.

We hope that our results will encourage experimentalists to develop new ways to determine solvation free energies of large molecules. Highly precise solvation free energies of amino acids and peptides not only might serve to validate our data, but may also be invaluable for the refinement of simulation methods and the theoretical understanding of the solvation contributions to all biomolecular processes.

SUPPORTING MATERIAL

Electrostatic and nonpolar contributions to the solvation free energies, implications for implicit solvent models, and references (81–83) are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)05119-3.

We thank Giorgos Archontis, David Mobley, and Chris Oostenbrink, who provided helpful comments on an early draft of this manuscript. We especially thank Bernard Brooks for providing us with the computational resources for the protein point mutation calculations on the LoBoS cluster.

This work was supported financially by project No. P-19100 of the Austrian Science Fund (Fonds zur Förderung der Wissenschaftlichen Forschung) and was also supported in part by the Intramural Research Program of the

National Heart, Lung, and Blood Institute of the National Institutes of Health.

REFERENCES

1. Benson, S. W. 1976. Thermochemical Kinetics: Methods for Estimation of Thermochemical Data and Rate Parameters, 2nd Ed. John Wiley, New York.
2. Lumry, R., and S. Rajender. 1970. Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers.* 9:1125–1227.
3. Merz, Jr., K. M. 2010. Limits of free energy computation for protein-ligand interactions. *J. Chem. Theory Comput.* 6:1018–1027.
4. Tanford, C. 1962. Contribution of hydrophobic interactions to stability of globular conformation of proteins. *J. Am. Chem. Soc.* 84:4240–4247.
5. Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.
6. Dyson, H. J., P. E. Wright, and H. A. Scheraga. 2006. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. USA.* 103:13057–13061.
7. Langhorst, U., J. Backmann, …, J. Steyaert. 2000. Analysis of a water-mediated protein-protein interaction within RNase T1. *Biochemistry.* 39:6586–6593.
8. Tarek, M., and D. Tobias. 1999. Environmental dependence of the dynamics of protein hydration water. *J. Am. Chem. Soc.* 121:9740.
9. Janin, J. 1999. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure.* 7:R277–R279.
10. Palomer, A., J. J. Pérez, …, D. Mauleón. 2000. Modeling cyclooxygenase inhibition. Implication of active site hydration on the selectivity of ketoprofen analogues. *J. Med. Chem.* 43:2280–2284.
11. Guthrie, J. P. 2009. A blind challenge for computational solvation free energies: introduction and overview. *J. Phys. Chem. B.* 113:4501–4507.
12. Nicholls, A., D. L. Mobley, …, V. S. Pande. 2008. Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. *J. Med. Chem.* 51:769–779.
13. Wolfenden, R. 1978. Interaction of the peptide bond with solvent water: a vapor phase analysis. *Biochemistry.* 17:201–204.
14. Wolfenden, R., L. Andersson, …, C. C. Southgate. 1981. Affinities of amino acid side chains for solvent water. *Biochemistry.* 20:849–855.
15. Biswas, K. M., D. R. DeVido, and J. G. Dorsey. 2003. Evaluation of methods for measuring amino acid hydrophobicities and interactions. *J. Chromatogr. A.* 1000:637–655.
16. Kyte, J., and R. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.
17. Nys, G. G., and R. F. Rekker. 1973. Statistical analysis of a series of partition-coefficients with special reference to predictability of folding of drug molecules—introduction of hydrophobic fragmental constants (*F*-values). *Chim. Ther.* 8:521–535.
18. Nys, G. G., and R. F. Rekker. 1974. Concept of hydrophobic fragmental constants (*F*-values). 2. Extension of its applicability to calculation of lipophilicities of aromatic and heteroaromatic structures. *Chim. Ther.* 9:361–375.
19. Hansch, C., and A. J. Leo. 1979. Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley Interscience, New York.
20. Mannhold, R., and R. F. Rekker. 2000. The hydrophobic fragmental constant approach for calculating log *P* in octanol/water and aliphatic hydrocarbon/water systems. *Perspect. Drug Discov. Des.* 18:1–18.
21. Cabani, S., P. Gianni, …, L. Lepori. 1981. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J. Solution Chem.* 10:563–595.
22. Karplus, P. A. 1997. Hydrophobicity regained. *Protein Sci.* 6:1302–1307.
23. Lazaridis, T., G. Archontis, and M. Karplus. 1995. Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. *Adv. Protein Chem.* 47:231–306.
24. Avbelj, F., and R. L. Baldwin. 2006. Limited validity of group additivity for the folding energetics of the peptide group. *Proteins.* 63:283–289.
25. Della Gatta, G., T. Usacheva, …, D. Ichim. 2006. Thermodynamics of solvation of some small peptides in water at $T = 298.15$ K. *J. Chem. Thermodyn.* 38:1054–1061.
26. Baldwin, R. L. 2007. Energetics of protein folding. *J. Mol. Biol.* 371:283–301.
27. Robertson, A. D., and K. P. Murphy. 1997. Protein structure and the energetics of protein stability. *Chem. Rev.* 97:1251–1268.
28. Ben-Naim, A., and Y. J. Marcus. 1984. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* 81:2016–2027.
29. Ben-Naim, A. 1992. Statistical Thermodynamics for Chemists and Biochemists. Plenum, New York.
30. Ben-Naim, A., and R. Mazo. 1997. Size dependence of solvation Gibbs energies: a critique and a rebuttal of some recent publications. *J. Phys. Chem. B.* 101:11221–11225.
31. Dill, K. A. 1997. Additivity principles in biochemistry. *J. Biol. Chem.* 272:701–704.
32. König, G., and S. Boresch. 2009. Hydration free energies of amino acids: why side chain analog data are not enough. *J. Phys. Chem. B.* 113:8967–8974.
33. Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.
34. Hermann, R. B. 1972. Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with cavity surface area. *J. Phys. Chem.* 76:2754–2759.
35. Chothia, C. 1974. Hydrophobic bonding and accessible surface area in proteins. *Nature.* 248:338–339.
36. Reynolds, J. A., D. B. Gilbert, and C. Tanford. 1974. Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc. Natl. Acad. Sci. USA.* 71:2925–2927.
37. Ooi, T., M. Oobatake, …, H. A. Scheraga. 1987. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. USA.* 84:3086–3090.
38. Wesson, L., and D. Eisenberg. 1992. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* 1:227–235.
39. Makhatadze, G. I., and P. L. Privalov. 1995. Energetics of protein structure. *Adv. Protein Chem.* 47:307–425.
40. Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.
41. Lazaridis, T., and M. Karplus. 1999. Effective energy function for proteins in solution. *Proteins.* 35:133–152.
42. Ferrara, P., J. Apostolakis, and A. Caflisch. 2002. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins.* 46:24–33.
43. Yunger, L. M., and R. D. Cramer, 3rd. 1981. Measurement of correlation of partition coefficients of polar amino acids. *Mol. Pharmacol.* 20:602–608.
44. Roseman, M. A. 1988. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* 200:513–522.
45. Chang, J., A. M. Lenhoff, and S. I. Sandler. 2007. Solvation free energy of amino acids and side-chain analogues. *J. Phys. Chem. B.* 111:2098–2106.
46. Shirts, M. R., J. W. Pitera, …, V. S. Pande. 2003. Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* 119:5740–5761.

47. Reif, M., P. Hünenberger, and C. Oostenbrink. 2012. New interaction parameters for charged amino acid side chains in the GROMOS force field. *J. Chem. Theory Comput.* http://dx.doi.org/10.1021/ct300156h.

48. Tembe, B. L., and J. A. McCammon. 1984. Ligand-receptor interactions. *Comput. Chem.* 8:281–283.

49. Brooks, B. R., C. L. Brooks, 3rd, …, M. Karplus. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.

50. Brooks, B. R., R. E. Bruccoleri, …, M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* 4:187–217.

51. MacKerell, Jr., A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of protein. *J. Phys. Chem. B.* 102:3586–3616.

52. MacKerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.

53. Bennett, C. H. 1976. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* 22:245–268.

54. Kirkwood, J. G. 1935. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* 3:300–313.

55. Jorgensen, W. L., H. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926.

56. Neria, E., S. Fischer, and M. Karplus. 1996. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105:1902.

57. Lamoureux, G., and B. Roux. 2003. Modeling induced polarization with classical Drude oscillators: theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* 119:3025–3039.

58. Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.* 31:1695–1697.

59. van Gunsteren, W. F., and H. J. C. Berendsen. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34:1311–1327.

60. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

61. König, G., and S. Boresch. 2010. Non-Boltzmann Sampling and Bennett's acceptance ratio method: how free energy simulations can profit from bending the rules. *J. Comput. Chem.* http://dx.doi.org/10.1002/jcc.21687 Early View.

62. Jo, S., T. Kim, …, W. Im. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29:1859–1865.

63. Villa, A., and A. E. Mark. 2002. Calculation of the free energy of solvation for neutral analogs of amino acid side chains. *J. Comput. Chem.* 23:548–553.

64. Deng, Y. Q., and B. Roux. 2004. Hydration of amino acid side chains: nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B.* 108:16567–16576.

65. Mobley, D. L., E. Dumont, …, K. A. Dill. 2007. Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B.* 111:2242–2254.

66. Reference deleted in proof.

67. Mobley, D. L., C. I. Bayly, …, K. A. Dill. 2009. Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput.* 5:350–358.

68. Staritzbichler, R., W. Gu, and V. Helms. 2005. Are solvation free energies of homogeneous helical peptides additive? *J. Phys. Chem. B.* 109:19000–19007.

69. Seeliger, D., and B. L. de Groot. 2010. Protein thermostability calculations using alchemical free energy simulations. *Biophys. J.* 98:2309–2316.

70. Veenstra, D. L., and P. A. Kollman. 1997. Modeling protein stability: a theoretical analysis of the stability of T4 lysozyme mutants. *Protein Eng.* 10:789–807.

71. Fernández-Escamilla, A. M., M. S. Cheung, …, L. Serrano. 2004. Solvation in protein folding analysis: combination of theoretical and experimental approaches. *Proc. Natl. Acad. Sci. USA.* 101:2834–2839.

72. Blaber, M., J. D. Lindstrom, …, B. W. Matthews. 1993. Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala→Ser and Val→Thr substitutions in T4 lysozyme. *Biochemistry.* 32:11363–11373.

73. Jiang, N., and C. Frieden. 1993. Intestinal fatty acid binding protein: characterization of mutant proteins containing inserted cysteine residues. *Biochemistry.* 32:11015–11021.

74. Dupureur, C. M., B. Z. Yu, …, M. D. Tsai. 1992. Phospholipase A2 engineering. The structural and functional roles of aromaticity and hydrophobicity in the conserved phenylalanine-22 and phenylalanine-106 aromatic sandwich. *Biochemistry.* 31:10576–10583.

75. Kumar, M. D. S., K. A. Bava, …, A. Sarai. 2006. ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res.* 34(Database issue):D204–D206.

76. Ponder, J. W., C. Wu, …, T. Head-Gordon. 2010. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B.* 114:2549–2564.

77. Shivakumar, D., J. Williams, …, W. Sherman. 2010. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* 6:1509–1519.

78. Wittgenstein, L. 1963. Logical Philosophical Treatise [Tractatus Logico-Philosophicus]. Suhrkamp, Berlin.

79. Feig, M., and C. L. Brooks, 3rd. 2004. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* 14:217–224.

80. Chen, J., C. L. Brooks, 3rd, and J. Khandogin. 2008. Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr. Opin. Struct. Biol.* 18:140–148.

81. Lee, M. S., M. Feig, …, C. L. Brooks, 3rd. 2003. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* 24:1348–1356.

82. Im, W., M. S. Lee, and C. L. Brooks, 3rd. 2003. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* 24:1691–1702.

83. Haberthür, U., and A. Caflisch. 2008. FACTS: Fast analytical continuum treatment of solvation. *J. Comput. Chem.* 29:701–715.